International Journal of Modeling and Optimization, Vol. 11, No. 1, February 2021

Community Safety and Well-being in Touristic Spots Using Open Data

Dineu Assis, Miguel de Castro Neto, and Marcel Motta

***Abstract*—There are many different reasons that can lead a tourist to decide which destination will be chosen on his/her next trip. Besides knowing what are the attractions that must be visited, it is also common to look for more information regarding the overall safety and well-being conditions of travel destinations. Usually shared by local authorities, this kind of information can also be found in a less structured form through public sources, such as web sites and social platforms. However, there are a couple of challenges to be considered: the predominance of unstructured data; the lack of a common standard to distinguish safe and unsafe places; the distinct period needed to update the collected data. In this study, the proposed model combines official census data with open data, social platforms and other online sources, allowing the definition of a score for touristic spots in Lisbon. The resulting score should be able to quantify the community safety and well-being, as well as to identify threats and opportunities for the local tourism industry. Furthermore, it would not only help tourists in their traveling decisions but also, allow decision-makers to track socioeconomic issues and to support public management through a data-driven approach.**

***Index Terms*—Community safety, well-being, tourism, smart cities, urban analytics, data mining.**

I. Introduction

The tourism industry in Portugal has grown steadily in recent years. Since 2010 the volume of international arrivals worldwide has increased by around 5%, every single year [1], [2], [3]. Europe concentrates more than half of this result and when the benefits of this growth are analyzed in the country aspect of Portugal they become relevant both in the financial context – through positive impacts on revenue [4] [5] – as well as in the planning of public policies, which need to be developed to accompany sustainable growth of tourism and its impact on the well-being of locals.

In addition to the rise of the public revenues, the development of the tourism sector brings with it concerns about crime rates in cities and tourist sites. A tourist can be described as is a sum of relationships and phenomena that result in travel and experiences as non-residents [6]. There are many reasons that can lead a tourist to choose the possible destinations for his/her trip, such as safety conditions, logistics and comfort. However, as these factors can be perceived subjectively, making an informed decision can be challenging, as some may find it difficult to interpret indices, recommendations or to gain access to local news sources.

Similarly, for the local community of a touristic destination, it is relevant to know what indicators are adequate to measure quality of life and tourism development. In this context, a diverse set of factors can become critical while determining areas of interest for developing tourism activities. This way public policies can be planned and implemented to improve these indicators and provide social good for the population. More than that, it is also relevant to know if public policies adopted for a city (or district) follow any kind of standards or have been monitored in order to improve quality of life.

To fulfill the objectives proposed in this study, open data initiatives, considered as defining elements of emerging smart cities, as they provide citizens with the tools necessary to create new, innovative services or applications [7] play a critical role.

II. Literature Review

The literature starts defining "safe places" as a broad concept of community: the earliest publication referring to the term "Safe Community" still refers to the situation analyzed in the context of England in 1986 [8]. At that time, the British government underwent a change of concept. from "crime prevention" to "safety in communities". The goal with this change was to expand the responsibility of crime prevention beyond the police and to consider social aspects of crime that are affected by perceived risk, organizations, families and individuals. As a result of this change, the government leaderships of England and Wales ended up describing the concept of "Safe Communities" as follows: "Safe Community is generally one or more community actions to inhibit and remedy the causes and consequences of criminal, intimidating and related antisocial behavior. Its goal is to ensure sustainable reductions in crime as well as the perception of crime in local communities. Its approach is based on forming multi-agency partnerships between the public, private and voluntary sectors to formulate and introduce community crime measures".

Complementarily to this definition, the "well-being" state is achieved when all the psychological, social and physical resources required by the community are used to meet a particular psychological, social and/or physical challenge [9]. When individuals have more challenges than resources, the balance falls apart and the welfare state changes.

However, in order to evaluate Community Safety and Well-Being (CSWB) – even though its measurement is abstract, as already mentioned – there is a need to create an index that fulfils the role of evaluating the contribution of

Manuscript received June 30, 2020; revised August 23, 2020. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

D. Assis, M. C. Neto, the M. Motta are with NOVA Information Management School, Universidade Nova de Lisboa. Lisbon, Portugal (e-mail: m20170164@novaims.unl.pt, mneto@novaims.unl.pt, mmotta@novaims.unl.pt).

DOI: 10.7763/IJMO.2021.V11.770

social in multisector to create conditions for global improvement. These indicators should aggregate the results of related sectors based on their results that should be shared [10], [11], [8]. When aggregated, there is an index that determines the CSWB level. By doing this, it will be possible to achieve scalability, complexity and systemic perspectives [12] – aiming to derive a result that can be associated to the whole community.

Typically, the results described during the process come from Economic, Health, Safety, Social and Environmental perspectives, that are going to be detailed in the next sections.

A. Social Perspective

The social dimension of "well-being" is understood by the classic components of "social capital" and its opportunities for access to public services. “Social capital” means divisions and identities by class, gender, religion, ethnicity, age, among others. Social conflict is considered as a main agent for well-being and collective political actions as possible solutions for handling with conflict.

Considering each of the suggested dimensions, the main indicators used to measure the proposed CSWB index for a given community are reported in [13]. In regard to the Social perspective, one of the items that was evaluated and that will also represent this perspective is the access to public services and facilities. This might extend to subjective characteristics, such as the way people perceive the quality of public services – whether fair or not – as well as the individual perception of the public structure and the efficiency of this sector.

Additionally, the relevance of the social context in communities is assessed through the sanitation conditions to which the population is subject [14]. The relevance of this issue is understood to be a systemic factor: sanitation is a factor that can be identified in isolation – as it may be restricted to a specific geographical area – however, sanitation issues will typically cause repercussions on the entire community, precisely due to social relations, which define the “social capital”.

Therefore, to rate touristic spots in the Social perspective, the availability of public buildings and issues raised in relation to the sanitation conditions will be considered.

B. Safety Perspective: Crime Modelling

Articles usually explore the occurrence of crime in events and unique occasions around the world (i.e. a concert, a sporting event, a conference), which allows the definition of crime and its related entities (author, victim, types of crimes, scene) [15]. In a specific case, where criminal occurrences were found in Auckland (New Zealand), it is possible to see variations in the volume of certain types of crime during the late 1990s and early 2000s. Additionally, it is also suggested a way to follow the evolution of the occurrences over time, corroborating the definition of a score at an opportune moment [15].

Following the same reference, it is clear that there is a concern to distinguish the place where the crime occurred. In Fig. 1, it is possible to check the correlation between the volume of assets lost for every type of crime scene. The data portrayed Auckland during a sporting event in the year 2000.

The proposed segmentation meets the normalization criteria that may compose a crime index score. But one should still explore not only the sites, but also the regions of the city and, to an even greater extent, its district.

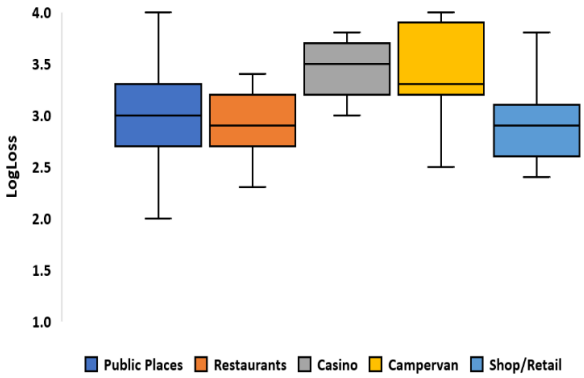

Fig. 1. Loss of assets in relation to the crime scene.

As with the Social perspective previously defined, the information related to the occurrence of the crimes (place where it was committed and the type of crime) will be responsible for composing the Safety perspective.

C. Economic Perspective: Hotels and Public Transport

Public transport networks are very critical components of large cities. It is difficult to estimate the capacity of these networks, especially in places where tourism directly is a great influencer of people who need to be served. Previous works suggests real-time analysis to anticipate actions and ensure greater resilience in public transport networks [16]. There is also a specific study on subway networks (which usually require greater investment) that considers adaptations to this transport system to ensure their availability given a planned schedule, but also with consequent action analysis of the facts in real time [17].

The concept of the term “public transport” is still being transformed over the years. In fact, after the emergence of “Smart Cities”, transport is no longer just a service offering that provides mobility to the people, but also considers the availability of information. It is no longer just a question of what means are available for transport from one place to another. People need to know when the transport will be available, what is the best available route, what is the best mode of transport to that destination, and the capacity of the available mode. In this context, previous works discussed solutions adopted in Budapest to improve the available vehicle traffic system as well as future implementations being made [18].

Based on the above sources, it is essential to use public data – preferably in real time – on the conditions and availability of the flow that buses, subways and any other public means of transport for a correct assessment of citizens' perception of a locality. city specific.

Still in the Economic context, in addition to public transport and its availability, there is also the figure of tourism and its derivates: hotels. The literature explores several possibilities of defining tourism and its relations with other sectors. There is also the concern to define the main actor that moves tourism, that is, the tourist himself.

Moving forward, the Economic perspective mentioned in this study will be described measuring transport and hotels availability, as well as their proximity to touristic spots, which will eventually measure the ability of each spot on

attracting new tourists.

D. Environment Perspective

Other indicators that represents the citizen's quality of life and are essential for the creation of a CSWB index correspond to the environmental quality indices. The availability of parks and green areas notoriously makes up the environmental picture of urban sub regions and measures the quality of life of these places. Miscellaneous occurrences or requests for public spaces that require government action (e.g. sanitation, maintenance of pavements, various situations related to public lighting or high noise levels, maintenance of green areas) define the role of the state regarding its obligations in managing the public environment and, consequently, its impact in people's daily life.

Health and Environment should be analyzed separately when observed in the context of the community [8]. Although they are intrinsically related within the community (the reduction of green areas – defined in the "Environment" perspective, for example, leads to variations in sanitation indicators – which belong to the "Health" perspective), the indicators will be treated independently.

To define what should be observed and thus guarantee the citizen a good quality of life in the Smart Cities, a survey was conducted in 2018 [19]. The term "Urban Sustainability" encompasses the perception of the population from the perspective of pollution indicators (sound/visual/climate), government actions, sustainable development and economy. All indicators cited in the study are linked to the perception of the quality of life of citizens living in the areas where the survey was conducted.

Similarly, the different forms of requests made by the population to governmental bodies to maintain the quality of life and to implement the concept of well-being in safe communities will be observed.

E. Health Perspective

It has been previously shown that one of the ways to model health quality in a community occurs is by assessing the availability of different public health modalities (e.g. health centers, hospitals, gyms) [13]. In this regard, the offer of Health Clinics near the observed points indicates that the quality of life of that place is relevant.

Comparing this with the "Environment" perspective, the health indicators will measure the availability of health services in areas close to tourist locations. Requests made by the population to governmental bodies for health maintenance and urban hygiene related to well-being in safe communities will also be observed.

III. Conceptual Model

In order to aggregate and evaluate all sources of information that define the perspectives described above, we chose to use a crawler-based model, defined as "a set of programs that are able to download pages interactively or automatically by extracting HTML content from predefined URLs [20]. A web crawler, for example, can be fed with a URL and then download all the content of the pages that are related to it in the form of hyperlinks.

In web crawler programs, it is possible to add calculations during searches to identify content that is considered most relevant to the search, or reject it (if it turns out to be duplicate or already visited content, for example). An important role for crawlers is to support search engines (e.g. Microsoft Bing, Google Search) in setting indexes after retrieving web page content. Regarding data mining, it is possible to build a crawler in a separate application and then perform the analysis of the data that is captured.

Using crawlers, all the essential information for defining community safety indices will be captured from a variety of sources related to previous topics available on the Internet. After collecting, treating and classifying the data, a model will be created to compute a score for each environment classified as "Tourist Attraction", assigning it a specific rating according to the evaluation of the previous assumptions.

In summary, the captured data will be associated with a category that will quantify the CSWB index in the 5 main perspectives:

- Safety: Crime and Police Action
- Social: Sanitation Activities, Availability of Public Buildings
- Health: Availability of Health Clinics and Urban Hygiene Conditions
- Economy: Tourism and Public Transportation
- Environment: Availability of Green Areas and Parks

As representative of the Economy perspective, also responsible for guiding the studies regarding the tourist aspect, we will define indicators related to the economy in Tourism and Public Transport. In the Social perspective, we will define indicators that evaluate the availability of public services and people's quality of life (i.e. sanitation conditions). For the Health perspective, the availability of health clinics and issues to be resolved within urban hygiene will be observed. Regarding the Safety perspective, we will observe the availability of police stations and indicators of urban crime. Finally, the Environment perspective will be defined by indicators related to visual and noise pollution, as well as the availability of green areas.

A. Planned Steps

First, it will be needed to define all data sources that will initially compose the proposed model. Then, for each source, the available subsections will be derived, which will relate to the final category that will compose the CSWB index, with a positive or negative value.

The CSWB index will be calculated by using the weighted average of the relative universe of the positive items that were found. The normalized score will be between 0 (zero) and 5 (five), where 0 (zero) will represent the minimum value and 5 (five) will represent the maximum value of the indicator.

In the end, the higher the value is, the better is also the result for that tourist attraction.

B. Data Normalization

For every perspective described, a rating must be assigned to tell whether the occurrence found has a positive or negative influence on each mapped context:

- Safety: Availability of police stations are positive items. Enhancement requests in any categories that

match public safety are negative. For evaluation purposes, only enhancement requests that are running at the time of data processing will be considered.

- Social: The proximity of public buildings of the observed locality will have a positive representation. Requests for improvement regarding the sanitation conditions will negatively influence the social result of that area.
- Health: The offer of Health Clinics in the observed area will have a positive representation. On the other hand, requests for improvement regarding Urban Hygiene conditions will negatively influence the social outcome of that area.
- Economy: Availability of hotels, bus and metro stops (public transport) will represent this perspective positively. Any requests regarding the poor quality of housing near the observed location will be negatively represented.
- Environment: The availability of green areas and parks near the observed area will have a positive representation. However, any requests for improvements in infrastructure conditions such as paving, road signs, street lighting as well as the maintenance of the green areas found will be assessed negatively.

In all cases where the population requests for the improvement of some service or public facility, only those requests that are being executed/under analysis at the time of data processing will be considered.

C. Process Flow

The following steps define the how data will be processed on each part or the pipeline.

Once the data sources have been defined, data will be extracted by crawling raw information from each source available. The crawler will parse HTML/Javascript and store the data in a more structured format.

In the second step the raw information obtained in the first step will be cleaned. At this time, the goal is to identify and treat anomalies found in the data set, such as erroneous values, missing data, among other data quality issues.

Considering that the data is structured and all relevant information was parsed, it will be needed to apply some filters in the result obtained from previous steps. These filters will help to identify patterns in the remaining texts and assign them a corresponding perspective. Items that could not find a proper perspective at this time will be stored in a staging area that might be used in an eventual manual assignment later.

Now that all data is ready to be read, data will be loaded into a Data Warehouse, allowing the definition of result indices and results comparison over time. Data Warehouses are multidimensional database structures that provide a single consistent source of management information for reporting and analysis [21].

The final step is to extract valuable information from the Data Warehouse, leading into a Key Performance Indicator (KPI) that will provide a business perspective about the CSWB indices.

Fig. 2 illustrates the workflow of the conceptual model previously described:

Fig. 2. Process flow.

D. KPI Definition

As formulated in the literature [8], the Safe Community and Welfare indicator requires the aggregation of different results observed by the analyzed perspectives, but no method is defined to be used for such aggregation. Thus, it was decided to aggregate these results by means of a weighted average [22]. The weights attributed to each of the perspectives will be responsible for quantifying their relevance in the CSWB index.

Given each perspective (n), the average score that will rank the results found will be defined in:

$$\frac{POS_n}{TOT_n} * W_n \tag{1}$$

Where POS represents all positive occurrences for that perspective, TOT represents all the occurrences for that same perspective, and W corresponds to the weight that will be used to evaluate the perspective relevance.

We have the final formula that represents the CSWB index by:

$$\frac{\sum_{Perspective=1}^{n} \frac{POS_n}{TOT_n} * W_n}{\sum_{Perspective=1}^{n} W_n} \tag{2}$$

By calculating the CSWB index, each perspective can be evaluated individually and each result can be aggregated and weighted according to how relevant that information is characterized in the context of the Safe Community and Welfare indicator.

Each equation represents a proportion of the occurrences found and classified in the observed perspective. These ratios are defined by the total positive occurrences (given by the factors prefixed with the label “POS”) over the total occurrences observed for this same perspective over the period analyzed (factors prefixed with the label “TOT”).

As already mentioned in previous topics, there is also the element that defines the “weight.” This is comprised of the factors prefixed with the label “W” and is responsible for giving due relevance to the analyzed perspective.

IV. Development

This section will detail the resources used, as well as the data structures that manage the results of the proposed model. Thus, the auxiliary interpretation of the data set can be performed from the sources described below, by adopting a classic Business Intelligence framework.

A. Extract-Transform-Load (ETL)

The input data is retrieved from operational data sources, which are designed for Online Transaction Processing (OLTP) systems. For this data to be effectively used in analytical tasks, it requires several procedures until it reaches an Online Analytical Processing (OLAP) system.

Moreover, this complete workflow consists of handling data with extract, transform, and load (ETL - Extract, Transform, Load) procedures, so it can be stored in the Data Warehouse. In contrast to typical transactional databases (i.e. OLTP), OLAP structures – such as the Data Warehouse – are more adequate for analytical purposes as they are able to reduce complexity, while maximizing integrity and efficiency.

As a result of this workflow, all critical information should be available in a data model, where it can be retrieved for subsequent analysis and to support business decisions. In order to store and manage data sources, database tables were used to store different data types for each source. Whenever the database is refreshed, incremental data is loaded and added for future analysis through the Data Warehouse.

B. Data Sources

The proposed model was based on data extracted as of August 2019. The following sources describe each one of the twelve data sources defined:

TABLE I: DATA SOURCES

Data Sources Name	Source
BaseNationalMonument	Lisbon City Hall
BasePublicMonument	Lisbon City Hall
BaseGreenArea	Lisbon City Hall
BasePark	Lisbon City Hall
BaseNeighborhood	Dados Abertos Website
BaseSubway	Lisbon City Hall
BaseBusRoute	Dados Abertos Website
BaseTourismDevelopment	Portugal Tourism Website
BasePoliceDepartment	Lisbon City Hall
BaseCrime	Portugal National Statistics Institute
BaseHealthCenter	Lisbon City Hall
BaseAppLx	Mobile Application (Minha Rua Lx)

Each source is considered as an "Open Data Source", and it is refreshed dynamically by its own providers or users (considering the Mobile Application reflects population opinions around certain perspectives). Added to this, there are some considerations regarding the format that every source is described: each source will be found in JSON or pure HTML formats.

As an example, the data sources structured with JSON follow the syntax very similar to what is seen on Fig. 3:


```
"type":"FeatureCollection",
"features":[
{
"type":"Feature",
"properties":{
"OBJECTID":1,
"COD_SIG":"2105705006001021",
"IDTIPO":"999",
"COD_SIG_EDIF":"2105705006001",
"NOME":"38ª Esquadra",
"MORADA":"Rua Ricardo Ornelas Lote 378, R/C-A",
"TIPO_UNIDADE":"Esquadra de Polícia",
"TELEFONE":"-",
"EMAIL":"-",
"CODPOSTAL":"-",
"FONTE":"PSP",
"MORADA_RMOG":"Rua Ricardo Ornelas ",
"GlobalID":"c9475c99-47fa-458a-a458-7a523d689846"
},
"geometry":{
"type":"Point",
"coordinates":[
-9.124781943050156,
38.752016513493594]
}}]
```


Fig. 3. Example of a JSON data source (BasePoliceDepartment).

The JSON example from above describes how a Police Department is defined for the city of Lisbon. It is possible to have access to its name, address, and the geographical coordinates (latitude and longitude).

The "BaseCrime" database was taken from a raw website, and its format is an HTML table, as it can be seen in Fig. 4:

Geographic localization (NUTS - 2013) (1)	Registered crimes (No.) by the police authorities by Geographic localization (NUTS - 2013) and Category of crime; Annual (3)
	Data reference period (2)
	2018
	Category of crime
	Total
	No.
Portugal	333 223
Continente	314 312
Região Autónoma dos Açores	8 956
Região Autónoma da Madeira	6 311

Fig. 4. Example of a HTML data source (BaseCrime).

C. Data Warehouse

The Data Warehouse, designed to manage the CSWB index and related data, is structured into dimensions (defined by the presented data sources) and fact tables (defined by each perspective already presented – Economy, Health, Social, Safety and Environment), all consolidated in a Star Schema. A Star Schema is the basic building block used in dimensional modelling and consists of one large central table called the "fact table", and a number of smaller tables called "dimension tables" which radiate out from the central table [21]. Fact tables contains quantitative measurements while dimension tables provide the basis for aggregating these measurements. Using this architecture will allow to compare touristic spots quantitatively across different attributes, supporting the proposed scoring process.

Therefore, for every data source described in the previous section, it is created a corresponding dimension (except the "DimTime", which represents a time frame and it is not derived from any data source) with relevant data from source tables that will assist in categorizing information further ahead.

Every dimension will focus on representing and describing one specific data source. The entities (i.e. dimensions) created to represent the information about to be processed and the relation with the source which it is representing can be found in the table below:

TABLE II: DATA WAREHOUSE DIMENSIONS

Dimension Name	Source
DimAppLx	BaseAppLx
DimBusRoutes	BaseBusRoute
DimCrimes	BaseCrime
DimHealthCenters	BaseHealthCenter
DimHotels	BaseTourismDevelopment
DimNeighborhoods	BaseNeighborhood
DimParks	BasePark
DimPoliceDepartments	BasePoliceDepartment
DimSubways	BaseSubway
DimTouristSpots	BaseNationalMonument
DimTime	Not Applied

The relationship and summarization between the dimensions are defined for every fact table (where each fact Table represents one of the perspectives previously presented), as listed below:

- "FactTouristSpotsEconomy": Summarizes Public Transportation and Hotels availability, and the "Minha Rua Lx" results related to economic events.
- "FactTouristSpotsHealth": This summarizes Health Clinics availability and "Minha Rua Lx" results for health events.
- "FactTouristSpotsSocial": This summarizes social indicators from "Minha Rua Lx", and the availability of Public Monuments.
- "FactTouristSpotsSafety": This summarizes crime events, availability of police departments and "Minha Rua Lx" safety events.
- "FactTouristSpotsEnvironment": This summarizes environmental indicators from "Minha Rua Lx", and also the availability of green areas such as gardens and parks.

D. Key Performance Indicators (KPIs)

For each perspective described in the previous sections, there were defined KPIs to allow the possibility to measure and compare results between the touristic spots. These KPIs are represented in the table below and their evaluation is defined such as an "up arrow" (meaning that greater values represent best indicators) or a "down arrow" (meaning the opposite):

TABLE III: KEY PERFORMANCE INDICATORS

Perspective	Measure Name	Evaluation
Safety	# Crime Events	↑
	# Police Departments	↑
	# Noisy Public Events (Minha Rua Lx)	↑
Social	# Public Monuments	↑
	# Sanitation Events (Minha Rua Lx)	↑
Health	# Health Clinics	↑
	# Urban Hygiene Events (Minha Rua Lx)	↑
Economy	# Bus Stops	↑
	# Subway Stations	↑
	# Hotels	↑
	# Housing Events (Minha Rua Lx)	↑
Environment	# Green Areas	↑
	# Gardens and Parks	↑
	# Green Areas Events (Minha Rua Lx)	↑
	# Sidewalk and Accessibility Events (Minha Rua Lx)	↑
	# Municipal Assets Events (Minha Rua Lx)	↑
	# Road Events (Minha Rua Lx)	↑
	# Street Lights Events (Minha Rua Lx)	↑

The eighteen indicators used to describe each perspective around its CSWB context are balanced between dynamic values – such as the results provided by the Mobile Application – and other typical static values – such as the annual reports provided by the Instituto Nacional de Estatística – Portugal's official authority for statistical studies, responsible for carrying out public enquiries and collecting/providing data at country level.

V. RESULTS AND DISCUSSION

Every national monument defined in the "DimTouristSpots" (63 national monuments extracted from the original "BaseNationalMonument") was rated by using the indicators defined in the previous section and after all these values were evaluated, the final result was represented by a scale from 0 to 5.

In order to give a better view about the final result, the following chart represents the consolidated result for each touristic spot following the "proximity axis" [8]. That compares each touristic spot starting from an "ideal entity" – an ideal touristic spot – compounded by the best results from each spot around every perspective defined previously, followed by the real results found during the data analysis.

In the end, all the public monuments were listed and their score indices ranked to allow a comparison across every perspective described along this study.

In order to make it easier to visualize the results, it was decided to split the values into quartiles: a quartile can help to decide which group will require more or less attention when it comes the time to create policies that will bring impact to the touristic experience. The results can be seen in the Fig. 5:

The "Score Q25" series groups the results 25% lower. These results are part of the first quartile of the database. The "Score Q50" series represents the results from Quartile 2 (those that are higher than the limit defined by Quartile 1 and lower than the one defined by Quartile 3). Finally, the results grouped in the "Score Q75" series are 25% higher.

A first approach shows which tourist sites need immediate response from the public sector: all those found in "Score Q25".

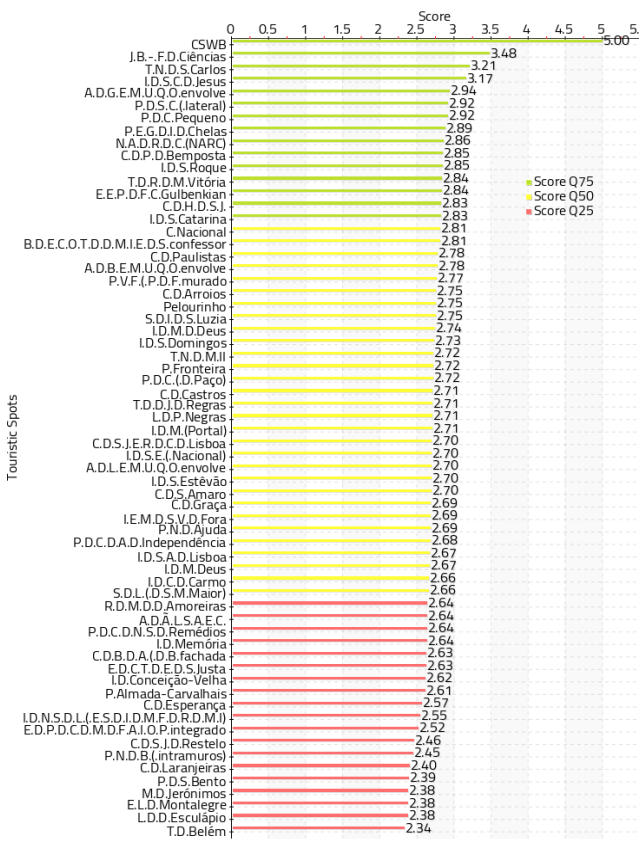

Fig. 5. Proximity axis.

Added to this, it is possible to evaluate eventual opportunities to improve the quality of life around the spots, superimposing the results of the equivalent perspective and comparing the index evaluated in each one of them. Fig. 6 shows how two spots can be compared over the five perspectives already defined:

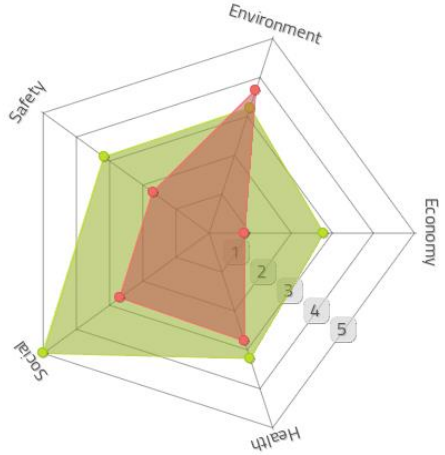

Fig. 6. Comparison between "Botanical Garden" (in green) and "Torre de Belém" (in orange).

As shown in Fig. 7, among the observed perspectives, touristic spots in Lisbon are defined by a higher score for the Environment perspective (mean at 0.70) and a significantly lower score for the Economy perspective (mean 0.25). Additionally, the Safety and Social perspectives have the highest dispersion (standard deviation at 0.10), while the Environment perspective has the lowest dispersion (standard deviation at 0.06). These parameters can give a better understanding of perspectives in which the city is thriving or struggling, as well as how these characteristics can impact the tourism industry. Overall, it suggests actions need to be taken for improving the local economy and to increase the consistency of social and safety indicators across the city.

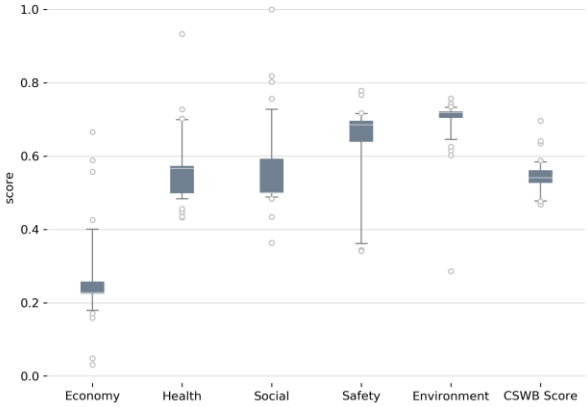

Fig. 7. Box plots for all observed perspectives.

Furthermore, the Pearson's correlation coefficient was calculated for each indicator to evaluate their level of dependence and to also identify the perspective that better explains the CSWB score given its covariance. As the CSWB score represents an aggregation of these perspectives, the correlation matrix should explain the linear relationship of each variable to the final score.

As shown in Fig. 8, no outstanding correlation is identified among the five perspectives. However, it was found that the Social perspective is the variable that better explains the CSWB score, as their correlation of 0.74 indicates a high positive correlation between them. Conversely, the Environment perspective seems to not explain the variance of the CSWB score, as their correlation of -0.05 is not significant. This behavior reinforces the need of adjusting the weights for each indicator to create a balanced overall score.

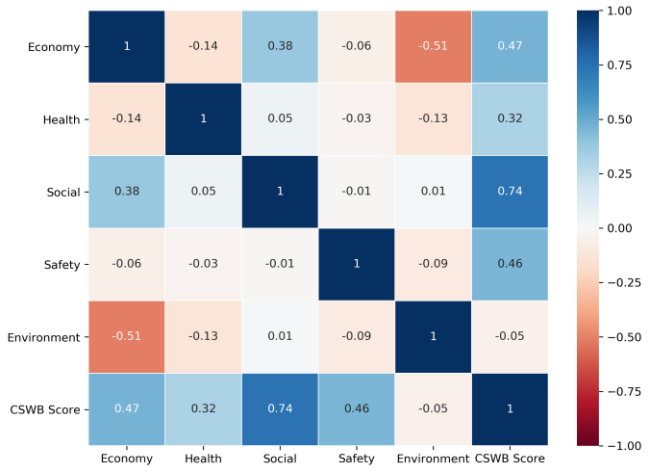

Fig. 8. Correlation matrix.

Moving forward, the the scores obtained from each perspective allowed the creation of clusters using Ward's hierarchical clustering method. This method allowed to objectively identify similarities among touristic spots in terms of the observed indicators, simplifying the data analysis process. Hence, the dataset was split into three separate groups, as detailed in Fig. 9:

- Red cluster: characterized by high Environment

indicators, but low Economy and Safety indicators. The lowest CSWB scores and the most distant spots to the city center belong to this cluster.

- Blue cluster: characterized by high Economy and Social indicators. The highest CSWB scores belong to this cluster.
- Yellow cluster: characterized by high Safety and Health indicators. The densest cluster, as most spots are concentrated on the city center.

Fig. 9. Hierarchical clustering results.

Considering the spatial data evaluated, the best spot found was the "Botanical Garden", which obtained the score of 3.48. At the other end is the iconic "Belém Tower", with a score of 2.34.

The chart area shown in the figure above allows us to evaluate that the result of the "Botanical Garden" is superior in almost all perspectives, being inferior only when it comes to the Environmental indicators.

The analysis around the Environment perspective shows the importance of dynamic indicator terms, in this case, provided by the application "Minha Rua Lx". It is easy to see that although both regions have a large number of green spaces and parks, it was in the evaluation of the entries found in the application that resulted in a better rating of the "Belém Tower".

Apparently, residents around the "Botanical Garden" are more concerned with reporting environmental problems and filing requests to the Lisbon City Hall than those residing near the "Belém Tower". In all other measures, the "Botanical Garden" always shows a better result, which justifies its position as the best rated touristic spot in Lisbon, as we can see below:

- Economy: High number of bus stops, subway stations, rooms and accommodation. In addition, a small number of open requests from residents in the mobile application. The difference in the availability of public transport near "Belém Tower" is evident, clearly shown by the data found in the KPI.
- Health: Almost equal to "Belém Tower", both have many entries in the mobile application, but the "Botanical Garden" still has a larger number of Health Clinics.
- Social: Because it is closer to other public monuments and has fewer entries in the mobile application, it is easy to see why the "Botanical Garden" was also highly ranked.

However, the granularity of the indicators available in the open data presented a challenge to overcome: for the "crime" indicator for example, the greatest detail available is only at the municipality level – which individually does not make sense, since all touristic spots are in the city of Lisbon. And, in order to overcome this granularity, it was decided to evaluate the crime conditions in the neighboring municipalities and to influence the results considering the geographical proximity of the points.

One item still to be solved refers to the behavior of the CSWB index throughout a longer period of time. Current data do not yet allow us to say whether/climate change can interfere with the interpretation of the information obtained here or the definition of the result throughout the year – which was also one of the objectives of this work. This is because current data sampling has not made it possible to compare them across other seasons.

Nonetheless, the robustness of the CSWB index was assessed using the same approach previously described but considering a posterior time frame. Using statistical methods, this new analysis compared the results obtained for August 2019 against the results for June 2020, to whether support or deny the reported results and conclusions.

However, by observing the differences between the two distributions, the Kolmogorov-Smirnov test rejects the null hypothesis that both samples are drawn from the same continuous distribution, with a 95% confidence level. More samples might be necessary to evaluate if the score distribution is consistent across the observed periods.

Fig. 10 shows the probability density distribution for the CSWB index in August 2019 and June 2020.

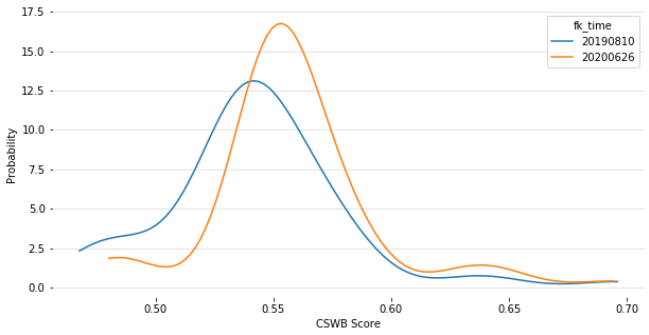

Fig. 10. Probability density distribution.

A comparison between the values generated in both samples can be found below. To make it easier to compare, we decided to emphasize the ten first results and the ten latest ones:

Starting with the comparison of the ten first results, we can see that even though the score index was slightly improved for the majority of the first 10 touristic spots, their final positions were affected by the improvement of other spots.

A similar behavior can be seen when the latest ten results are compared: in the table below, all the touristic sports had their final score improved. However, the biggest impact can

be seen when we follow the changes in the positions occupied by the touristic spots among the others. The stability of the score, with slightly improvements, suggests that the data evaluated is consistent and even though some of the sources have a dynamic component – such as the one that collect feedback from the population though a mobile app – the refreshed final score is close to the first analysis.

TABLE IV: COMPARISON BETWEEN 10 FIRST RESULTS

Touristic Spot	August 2019	June 2020	Position Change
Jardim Botânico - Faculdade de Ciências	3.48	3.47	0
Igreja do Sagrado Coração de Jesus	3.17	3.22	-1
Teatro Nacional de São Carlos	3.21	3.24	+1
Paços de São Cristóvão (Porta lateral)	2.92	2.92	-3
Padrão do Campo Pequeno	2.92	3.16	1
Edifício-Sede e Parque da Fundação Calouste Gulbenkian	2.84	3.15	1
Ascensor da Glória e meio urbano que o envolve	2.94	2.90	-4
Ascensor da Bica e meio urbano que o envolve	2.78	2.91	-2
Túmulo da Rainha D. Mariana Vitória	2.84	2.86	-4
Núcleo Arqueológico da Rua dos Correeiros (NARC)	2.86	2.82	-8

TABLE V: COMPARISON BETWEEN 10 FIRST RESULTS

Touristic Spot	August 2019	June 2020	Position Change
Igreja de Nª Sra. da Luz (Capela-Mor e Sepultura da Infanta D. Maria, Filha do Rei D. Manuel I)	2.55	2.71	+1
Edifício do Posto de Comando do Movimento das Forças Armadas, incluindo o património integrado	2.52	2.75	+15
Capela de São Jerónimo/Ermida do Restelo	2.46	2.80	+36
Palácio Nacional de Belém (conjunto intramuros)	2.45	2.83	+40
Cruzeiro das Laranjeiras	2.40	2.40	-5
Palácio de São Bento	2.39	2.48	0
Mosteiro dos Jerónimos	2.38	2.76	+29
Estátuas Lusitanas de Montalegre	2.38	2.76	+25
Lápide do Deus Esculápio	2.38	2.76	+25
Torre de Belém	2.34	2.69	+6

The evaluation of the touristic spots from the perspectives herein defined allows us to evaluate the opportunity for improvements in the touristic areas and thus provide a better quality of life not only for the local community residing in the evaluated region, as well as for the tourists who come to know Lisbon and Portugal. its surroundings. The comparison here between the first and last ranked in the rule to which it was applied shows that there are opportunities to direct investments and improve the status quo of touristic spots. Regarding the continuity of the model, it is noteworthy that the adaptation to new scenarios and challenges is feasible and may assist in decision making in other situations through the evaluation of indicators.

VI. CONCLUSION

The definition and measurement of community safety indicators [8], namely the Community Safety Well Being (CSWB) Index, allows to create a competitive ecosystem among the public entities responsible for maintaining touristic spots. Assessing and maintaining a "Safe Place" can no longer be viewed as reactive initiatives by a government or any other form of organization.

As shown, the definition of "Welfare" in cities is the responsibility of entities from various sectors who will need to work together for a favorable environment to be established. This ecosystem creates opportunities for improving social, economic, health, safety and environmental conditions around the points analyzed – each of which will be responsible for proposing actions to improve the individual results of the element in question.

The main objective of this article – the definition of an interactive model that evaluates tourist attractions from different perspectives responsible for characterizing the safety and welfare issues of the communities – was achieved. Even so, the model is flexible enough to be adapted to every new scenario. Regarding the definition of indicators in each of the proposed perspectives, the current solution cannot be considered as an exact science: indicators were selected from literature recommendations [13]. However, it should be noted that the selected indicators allow for certain adaptations – as long as the scope is not changed – following the framework defined in the literature [8].

The main question that surrounds this article ("how safe is this place?") can be answered by looking at the tourism data, since the analysis focused on each of the touristic spots defined by the Lisbon City Hall – and created a result for each of these. Therefore, depending on the area in which the tourist will want to visit, the model will support the decision making process by presenting an index of the most favorable places in all aspects: economic, social, health, safety and environment.

By capturing data and structuring it in the form of a Data Warehouse we were able not only to map the areas of interest (i.e. touristic spots with their geographic location and other information that is available from open data sources) and the characteristics that make up the surroundings of the touristic spots, but also create new information from the joint evaluation of these data.

The challenge to map and identify problems faced by tourists, was made possible by reading the open data provided by the Lisbon City Hall (such as the availability of public transport and the volume of hotels and rooms available, among others) but also by reading the open/dynamic data available through the "Minha Rua Lx" application.

Suggestions for future work associated with the proposed model:

- Adapt the model to allow using other entities such as local businesses (e.g. shopping centers, restaurants, etc) and other services/facilities.
- Analyze spatial relationships among touristic spots

and establish an influence area for each touristic spot.

- Adapt the model to other cities/countries, in particular where tourism is a major economic driver, to quantify the local CSWB scores and identify areas for improvement.
- Expand the observed time frame and formulate methods to evaluate the statistical significance for the estimated indicators.

Finally, the indicators evaluated along this work could help to gradually increase the number of visits to these sites, as they will become more attractive in all aspects. To achieve this, the indicators must be used to allow actions that could improve the current results for each touristic spot and hopefully change the tourist experience around the city of Lisbon.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

The research begun with Dineu Assis that managed to collect and analyze preliminary data based on the open data sources identified. Marcel Motta collaborated in the analysis by adding new perspectives and complementing results. All the research was leaded by Miguel de Castro Neto, who guided the group through the process with a special focus on the relevance of Smart Cities and the use of Open Data to improve urban spaces panning and management. All authors had approved the final version

Acknowledgment

We thank Universidade Nova de Lisboa for the support and opportunity to research on this Smart Cities subject and allow the possibility to improve life quality around the word through research. We thank the Camara Municipal de Lisboa (Lisbon City Hall) for providing relevant sets of data through its “Open Data Website” and “Minha Rua Lx” mobile application.

References

[1] “Estatísticas do turismo 2014,” Instituto Nacional de Estatística, Lisbon, Portugal, 2015.
[2] “Estatísticas do turismo 2015,” Instituto Nacional de Estatística, Lisbon, Portugal, 2016.
[3] “Estatísticas do turismo 2018,” Instituto Nacional de Estatística, Lisbon, Portugal, 2019.
[4] Portugal Recebeu mais de 60 milhões de turistas em 2016. Diário de Notícias. https://www.dn.pt/dinheiro/interior/portugal-recebeu-mais-de-60-milhoes-de-turistas-em-2016-5624832.html
[5] Turismo é o setor com maior crescimento no mundo. Dinheiro Vivo. [Online]. Available: https://www.dinheirovivo.pt/economia/turismo-e-o-setor-com-maior-crescimento-no-mundo
[6] N. Vanhove, *The Economics of Tourism Destinations*, Oxford, UK: Butterworth-Heinemann, 2005, p. 2.
[7] M. Neto, “Urban intelligence,” in *Proc. 18th Conferência da Associação Portuguesa de Sistemas de Informação (CAPSI)*, Santarém, Portugal, October 2018, pp. 12–13.
[8] C. Nilson, “Community safety and well-being: Concept, practice, and alignment,” *Journal of Community Safety and Well-Being*, vol. 3, no. 3, October 2018, pp. 96–104.
[9] R. Dodge, A. Daly, J. Huyton, and L. Sanders, “The challenge of well-being,” *International Journal of Well-Being*, vol. 2, no. 3, August 2012, pp. 222–235.
[10] A. Gilchrist, *The Well-Connected Community*, 2nd ed. Bristol, UK: The Policy Press, 2009.
[11] J. Kania and M. Kramer, “Collective impact,” *Stanford Social Innovation Review*, Winter 2011, pp. 36–41.
[12] S. Huddart. Seven years on and seven years out: Revisiting patterns, principles, and patterns in social innovation. https://thephilanthropist.ca/2017/04/seven-years-on-and-seven-years-out-revisiting-patterns-principles-and-practices-in-social-innovation
[13] P. Noursalehi and H. N. Koutsopoulos, “Real-time predictive analytics for improving public transportation systems’ resilience,” in *Proc. Bloomberg Data for Good Exchange Conference*, New York, USA, September 2016.
[14] M. J. Forjaz, M. Prieto-Flores, A. Ayala, C. Rodriguez-Blazquez, G. Fernandez-Mayoralas, F. Rojo-Perez, and P. Martinez-Martin, “Measurement properties of the Community Wellbeing Index in older adults,” *Quality of Life Research*, vol. 20, June 2011, pp. 733–743.
[15] M. Barker, S. Page, and D. Meyer, “Modeling tourism crime: The 2000 America's cup,” *Annals of Tourism Research*, vol. 29, no. 3, January 2002, pp. 762–782.
[16] H. Sung and R. G. Phillips, “Indicators and community well-being: Exploring a relational framework,” *International Journal of Community Well-Being*, vol. 1, October 2018, pp. 63–79.
[17] H. N. Koutsopoulos, P. Noursalehi, Y. Zhu, and N. Wilson, “Automated data in transit: Recent developments and applications,” in *Proc. 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, Naples, Italy, June 2017, pp. 604–609.
[18] J. Mezei and K. Lazányi, “Are we ready for smart transport? analysis of attitude towards public transport in budapest,” *Interdisciplinary Description of Complex Systems*, vol. 16, September 2018, pp. 369–375.
[19] D. Petrova-Antonova and S. Ilieva, “Smart cities evaluation – A survey of performance and sustainability indicators,” in *Proc. 2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, Prague, Czech Republic, August 2018, pp. 486–493.
[20] M. Thelwall, “A web crawler design for data mining,” *Journal of Information Science*, vol. 27, no. 5, October 2001, pp. 319–325.
[21] D. L. Moody and M. A. R. Kortink, “From enterprise models to dimensional models: A methodology for data warehouse and data mart design,” in *Proc. 2nd International Workshop on Design and Management of Data Warehouses (DMDW)*, Stockholm, Sweden, June 2000, pp. 1–12.
[22] A. Sharpe. (2004). Literature review of frameworks for macro-indicators. Centre for the Study of Living Standards, Ottawa, Canada. [Online]. Available: http://www.csls.ca/reports/LitRevMacro-indicators.pdf

Dineu Assis was born in Belo Horizonte, Brazil on July 1985 and is currently a master student at NOVA Information Management School from Universidade Nova de Lisboa, in Lisbon, Portugal. It has developed his academic work on business intelligence with the use of open data. Before this, back in 2008 he obtained the BSc. degree in computer science from the Universidade Federal de Minas Gerais.

He has been working for ICT (Information Technology and Communication) industries for 12 years, performing roles of supporting, project management, and reporting. 2018 he started working at Microsoft as a support escalation engineer in Portugal. He handles customer real problems related with solutions built on Microsoft business intelligence tools (integration Services, analysis services and reporting services).

Miguel de Castro Neto was born in Lisbon, Portugal in March 1967 and holds a PhD by Universidade Técnica de Lisboa on information systems applied to agronomy.

He is an associated dean at NOVA Information Management School, Lisboa, Portugal, where he created and coordinates NOVA Cidade – Urban Analytics Lab. He is also the Former Secretary of State for Spatial Planning and Nature Conservation in the Portuguese Government. He is a research interests include smart cities, urban analytics and business intelligence.

Prof. Castro Neto is the vice-president of the Data Science Portuguese Association, a member of the Portuguese Engineers Professional Association and of Sustainable Development Platform think tank. 217th Smart Cities Personality of the year by Green Business Week/Fundação AIP.

Marcel Motta is a PhD student at NOVA Information Management School from Universidade Nova de Lisboa, in Lisbon, Portugal. Currently he ddevelops his academic work on data science applied to urban analytics. In 2018 earned a MSc degree for information management, specialized on business intelligence and knowledge management, also awarded by NOVA Information Management School. Back in 2015 obtained a BSc degree on international relations from Universidade Estácio de Sá, in Rio de Janeiro, Brazil.

He worked for over 7 years in the IT/banking industries, where he performed roles in reporting, project management and risk management. Since 2019 works as a data scientist for Nova Cidade - Urban Analytics Lab, a research branch from NOVA Information Management School.